



University of Pennsylvania  
**ScholarlyCommons**

---

2017 ADRF Network Research Conference  
Presentations

ADRF Network Research Conference Presentations

---

11-2017

# Better Together: Three Models for Combining Public and Private Data to Create New Research Data Sets

Eddie Tzu-Yun Lin  
*University of Chicago*

Follow this and additional works at: [https://repository.upenn.edu/admindata\\_conferences\\_presentations\\_2017](https://repository.upenn.edu/admindata_conferences_presentations_2017)

---

Lin, Eddie Tzu-Yun, "Better Together: Three Models for Combining Public and Private Data to Create New Research Data Sets" (2017). *2017 ADRF Network Research Conference Presentations*. 9.  
[https://repository.upenn.edu/admindata\\_conferences\\_presentations\\_2017/9](https://repository.upenn.edu/admindata_conferences_presentations_2017/9)

This paper is posted at ScholarlyCommons. [https://repository.upenn.edu/admindata\\_conferences\\_presentations\\_2017/9](https://repository.upenn.edu/admindata_conferences_presentations_2017/9)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## Better Together: Three Models for Combining Public and Private Data to Create New Research Data Sets

# Better Together: Three Models for Combining Public and Private Data to Create New Research Data Sets

Eddie Tzu-Yun Lin  
tlin2@uchicago.edu

JPMorgan Chase  
Foundation

**Center for Data Science & Public Policy**



ALFRED P. SLOAN  
FOUNDATION



# Outline

- WDI Overview
- Challenges of Labor Market Data
- Open-Skills Project
- Combining Private and Public Sector Data
  - Use Public Sector Data for Validating Private Sector Data
  - Use Public Sector Data to Extract New Data from Private Sector Data
  - Use Private Sector Data for Creating New Public Sector Data
- What can you do today?

# The Workforce Data Initiative

## What We're Doing



# Open Data on Skills and Jobs

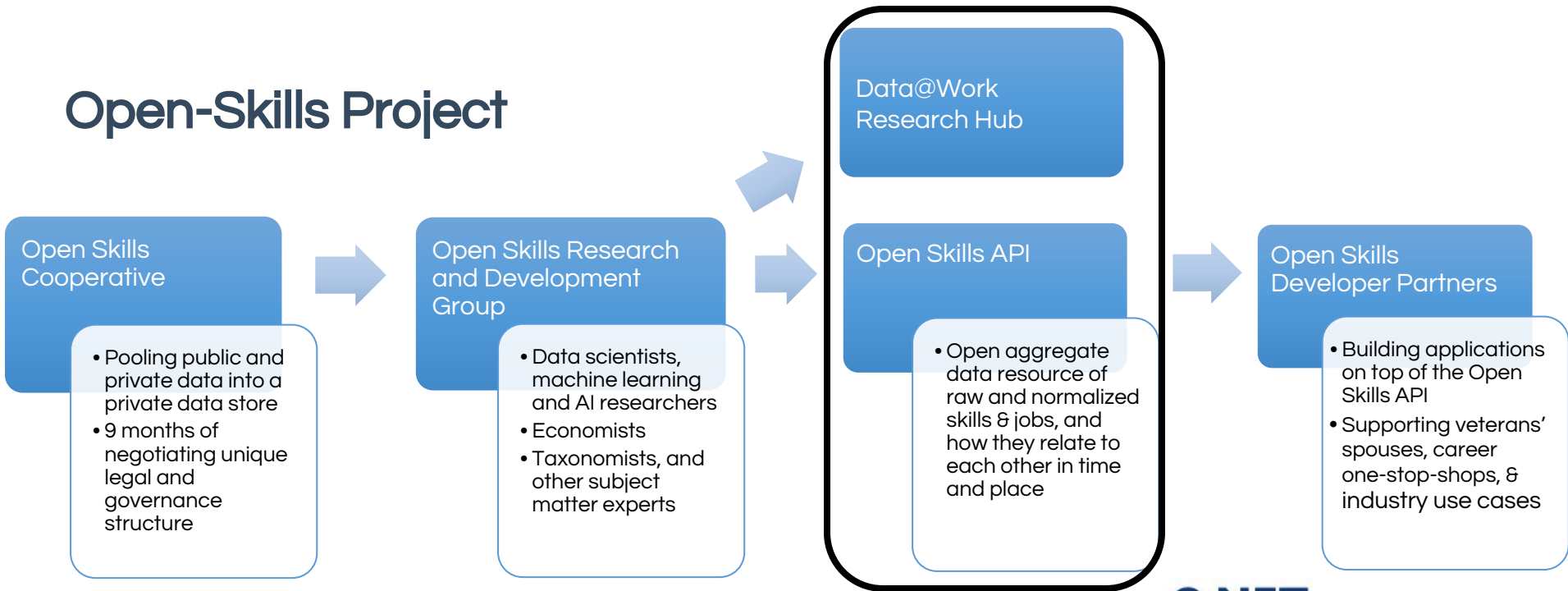
Opening  
Training  
Outcomes Data

# Supporting New Developer Resources and Third Party Tools

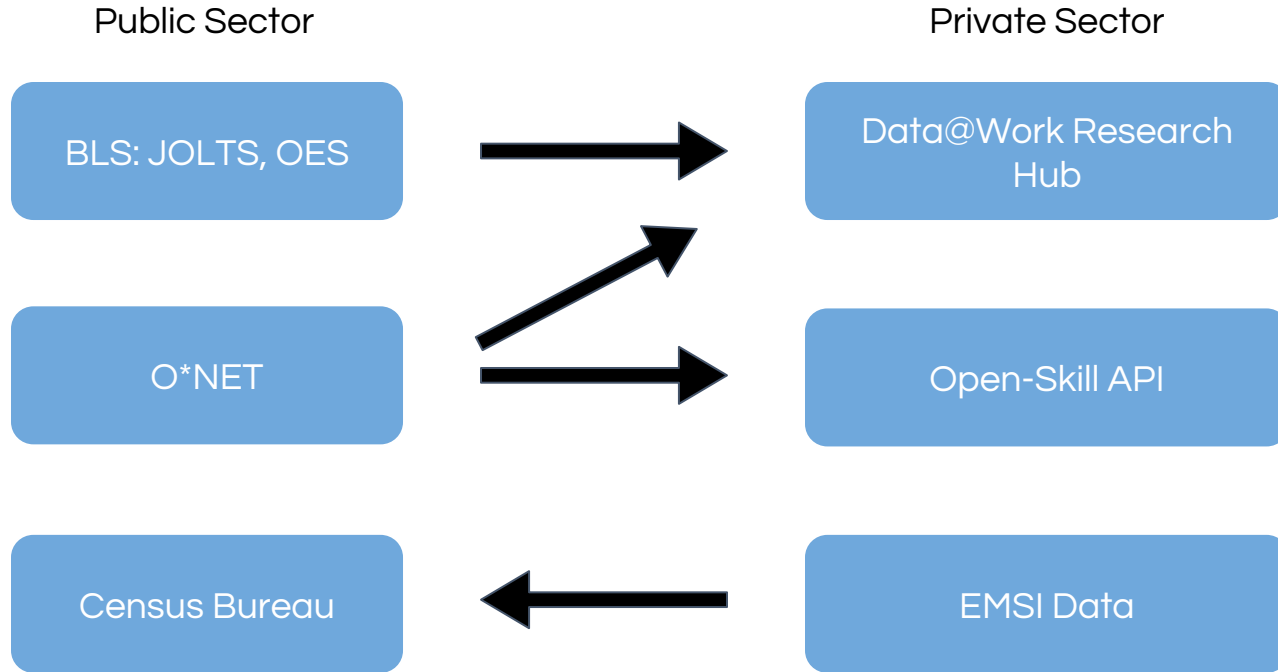
# Challenges of Labor Market Data

- Lack of geographic and and temporal granularity of survey-based labor market data sources
- Lack of intra-occupational specificity of national labor market taxonomies/ontologies
- High frictions to data access for more granular administrative data sources
- Lack of technical capacity among social science researchers for working with large, unstructured data sources
- Lack of transparency in methods for generated datasets from most Real-time Labor Market Information data vendors

# Open-Skills Project



# Three Models for Using Private and Public Sector Data





# Data@Work Research Hub Dataset

2013Q2.csv Cbsa Name	2013Q2.csv State Code	2013Q2.csv Title	2013Q2.csv Soc Code ...	2013Q2.csv Soc Code Com...	2013Q2.csv Onet Skills 1	2013Q2.csv Onet Skills 2	2013Q2.csv Onet Skills 3
Detroit-Warren-Dear...	MI	automotive production supervisor location relocat...	43-5053.00	1	rules	null	null
Hartford-West Hartf...	CT	dispatcher public safety	33-3051.01	1	transportation	null	null
Minneapolis-St. Paul...	MN	digital strategy director	11-2031.00	2	design	null	null
Philadelphia-Camden...	PA	database coordinator sql server	15-1141.00	1	programming	science	design
Auburn-Opelika, AL ...	AL	customer service representative	43-5071.00	2	guides	self	microsoft office
Bastrop, LA Micro Area	LA	pro	49-3023.02	2	transportation	auto	null
San Jose-Sunnyvale-...	CA	insurance agent opportunity	41-3021.00	1	auto	null	null
Sacramento--Rosevill...	CA	multiple positions available entry level manageme...	11-2022.00	5	route	player	sales and marketing
Chicago-Naperville-E...	IL	global business communications specialist	11-2021.00	1	linkedin	templates	facebook
Trenton, NJ Metro Ar...	NJ	administrator customer support clerk	41-2031.00	1	clerical	levels	null

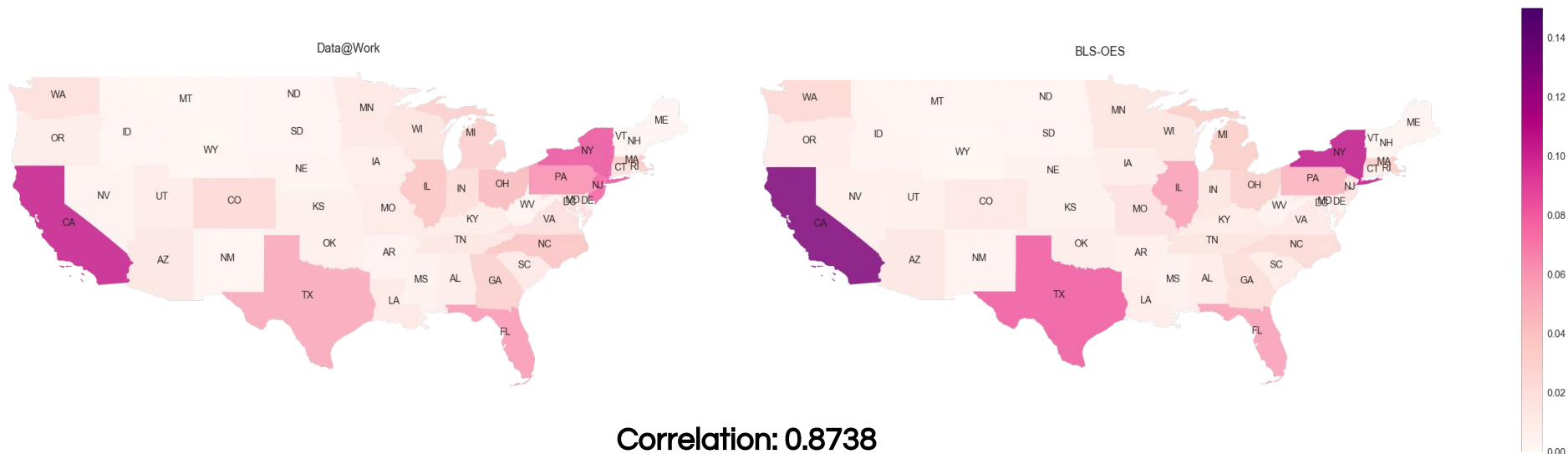
# Use Public Sector Data for Validating Private Sector Data

## Representativeness Analysis

Single Year Representativeness Summary Data Overview (2013)

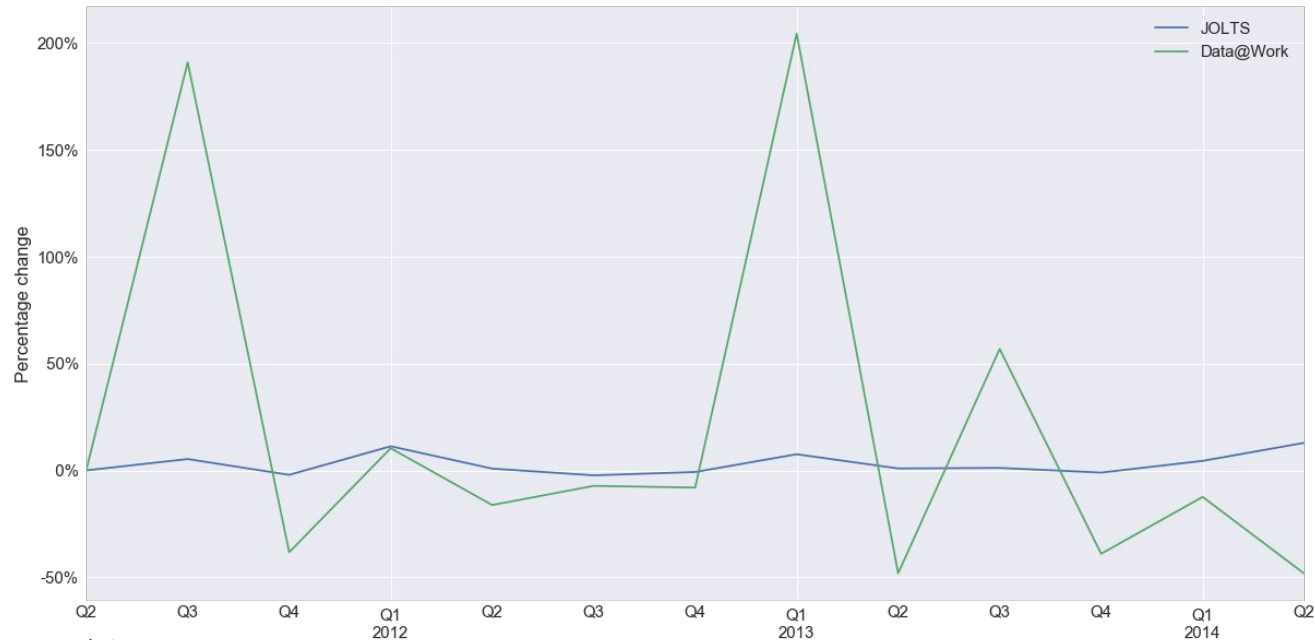
	OES	JOLTS	Data@Work
Type	Employment	Job opening	Job posting
Counts	169M	90M	2.4M
Geographic Granularities	586 MSAs	4 Regions	1004 CBSAs
Major Groups	22 (ONET)	26 (NAICS)	23 (ONET)

# Job Distribution by State: Data@Work vs. OES



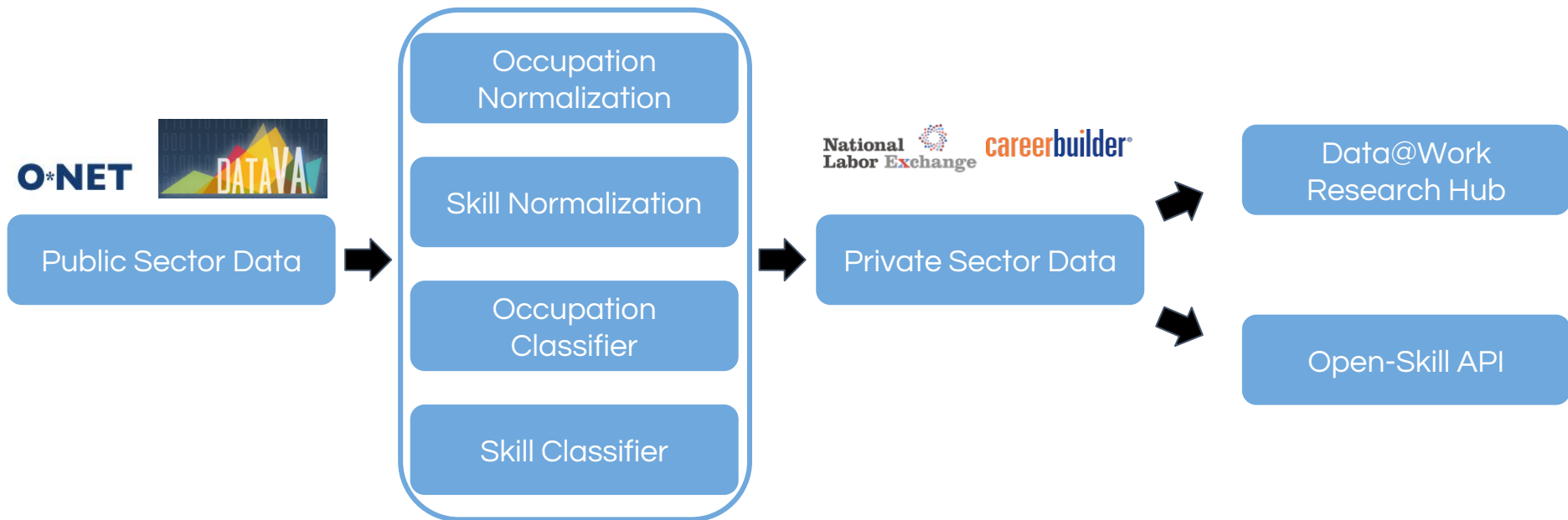
NOTE: Based on 2013 data. Does not include AK and HI.

# Percent Change in Data@Work Job Postings (contributed by CareerBuilder) vs. JOLTS Job Openings



NOTE: Based on 2011-2014 data.

# Use Public Sector Data to Extract New Data from Private Sector Data



# Use Private Sector Data for Creating New Public Data



Private Sector



DSaPP



Public Sector

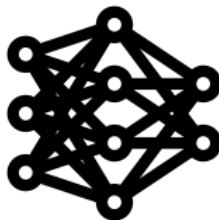


Open Data

Resume Data



Job Title Normalizer  
Occupation Classifier



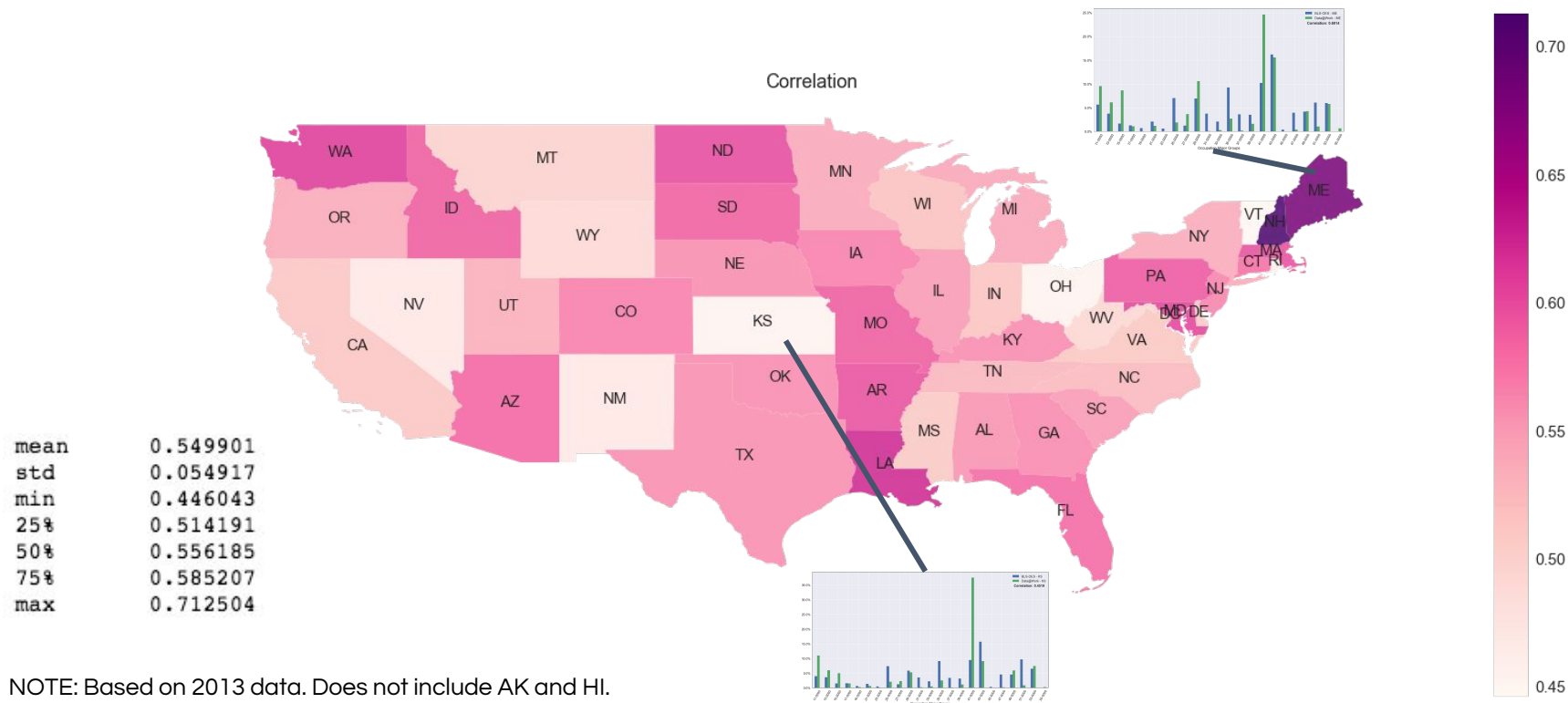
RSI Form



# What can you do today?

- **Sign up** to be part of the [Data@Work researcher community on Slack](#)
- **Explore the data** on the [Data@Work Research Hub](#)
- **Ask questions and provide suggestions** for additional datasets to the Data@Work on Slack

# Occupation Distribution Correlation by State



NOTE: Based on 2013 data. Does not include AK and HI.



# Use Public Sector Data to Extract New Data from Private Sector Data

## Open Skills API

A complete and standard data store for canonical and emerging skills, knowledge, abilities, tools, technologies, and how they relate to jobs.

Created by Work Data Initiative

See more at <http://www.dataatwork.org>

default

Show/Hide | List Operations | Expand Operations

GET	/jobs	Job Titles and Descriptions
GET	/jobs/{id}	Job Title and Description
GET	/jobs/{id}/related_skills	Skills Associated with a Job
GET	/jobs/{id}/related_jobs	Jobs Associated with a Job
GET	/jobs/autocomplete	Job Title Autocomplete
GET	/jobs/normalize	Job Title Normalization
GET	/jobs/unusual_titles	Unusual Job Titles
GET	/skills	Skill Names and Descriptions
GET	/skills/{id}	Skill Name and Description
GET	/skills/{id}/related_jobs	Jobs Associated with a Skill
GET	/skills/{id}/related_skills	Skills Associated with a Skill